Nº 12-593-XIF au catalogue

Guide de l'enseignant sur la découverte des données

Statistique Canada Statistics Canada

Canada

Comment obtenir d'autres renseignements

Toute demande de renseignements au sujet du présent produit ou au sujet de statistiques ou de services connexes doit être adressée à : Services aux clients, Division de la culture, tourisme et centre de la statistique sur l'éducation, Statistique Canada, Ottawa, Ontario, K1A 0T6 (téléphone : 613-951-7608; sans frais : 1-800-307-3382; télécopieur : 613-951-9040; courrier électronique : educationstats@statcan.ca).

Pour obtenir des renseignements sur l'ensemble des données de Statistique Canada qui sont disponibles, veuillez composer l'un des numéros sans frais suivants. Vous pouvez également communiquer avec nous par courriel ou visiter notre site Web à www.statcan.ca.

Service national de renseignements	1-800-263-1136
Service national d'appareils de télécommunications pour les malentendants	1-800-363-7629
Renseignements concernant le Programme des services de dépôt	1-800-700-1033
Télécopieur pour le Programme des services de dépôt	1-800-889-9734
Renseignements par courriel	infostats@statcan.ca
Site Web	www.statcan.ca

Renseignements pour accéder au produit

Le produit nº 12-593-XIF au catalogue est disponible gratuitement sous format électronique. Pour obtenir un exemplaire, il suffit de visiter notre site Web à www.statcan.ca et de choisir la rubrique Publications.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui doivent être observées par les employés lorsqu'ils offrent des services à la clientèle. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées dans le site www.statcan.ca sous À propos de nous > Offrir des services aux Canadiens.

Remerciements

Ce guide a été produit par le Programme de soutien à l'éducation de Statistique Canada sous la direction de Mary Townsend. L'équipe de rédaction dirigée par Elise Mennie tient à remercier Angela McCanny, une enseignante d'Ottawa au primaire, qui a entrepris l'élaboration de ce guide afin de répondre à un besoin exprimé par ses pairs. Elle a contribué à la réalisation du projet grâce à son expertise et y a consacré beaucoup de temps.

Message spécial aux enseignants

Avec l'aide du milieu de l'enseignement, Statistique Canada a conçu une vaste gamme d'activités, de plans de leçons et de trousses pédagogiques liés aux programmes d'études de la plupart des matières scolaires. Les enseignants utilisent ces ressources pour intégrer des informations canadiennes d'actualité dans leur enseignement et pour accroître l'« alphabétisme » statistique des élèves. Consultez la section « Ressources éducatives » du site Web de Statistique Canada à www.statcan.ca.

2008-03305/

Statistique Canada

Division de la culture, tourisme et centre de la statistique sur l'éducation

Guide de l'enseignant sur la découverte des données

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2007

Tous droits réservés. Le contenu de la présente publication électronique peut être reproduit en tout ou en partie, et par quelque moyen que ce soit, sans autre permission de Statistique Canada, sous réserve que la reproduction soit effectuée uniquement à des fins d'étude privée, de recherche, de critique, de compte rendu ou en vue d'en préparer un résumé destiné aux journaux et/ou à des fins non commerciales. Statistique Canada doit être cité comme suit : Source (ou « Adapté de », s'il y a lieu) : Statistique Canada, année de publication, nom du produit, numéro au catalogue, volume et numéro, période de référence et page(s). Autrement, il est interdit de reproduire le contenu de la présente publication, ou de l'emmagasiner dans un système d'extraction, ou de le transmettre sous quelque forme ou par quelque moyen que ce soit, reproduction électronique, mécanique, photographique, pour quelque fin que ce soit, sans l'autorisation écrite préalable des Services d'octroi de licences, Division des services à la clientèle, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Juin 2007

N° 12-593-XIF au catalogue
ISBN 978-0-662-73193-1

Périodicité : hors série

Ottawa

This publication is available in English upon request (catalogue no. 12-593-XIE).

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Table des matières

Introduction 5
Des compétences essentielles pour les élèves canadiens 5
La recherche statistique 5

1. Choisir l'ensemble de données 6
Données primaires 6
Qu'est-ce qu'on entend par données primaires? 6
Ressources pour la collecte de données primaires 6
Recensement à l'école 6
Les statistiques : le pouvoir des données! 7
Enquête sur les ménages et l'environnement — édition scolaire 7
Données secondaires 7
Qu'est-ce qu'on entend par données secondaires? 7
Sujets pour lesquels les données secondaires s'avèrent utiles 7
Sources de données secondaires canadiennes 8
E-STAT 8
Profils des communautés 8
Tableaux sommaires 8
Indicateurs de la santé 8
Données du recensement 8
Le Quotidien 9
Publications 9
Fichiers de microdonnées 9
Sources de données secondaires internationales 9
Données internationales du Recensement à l'école 9
Données internationales de E-STAT 9
Sites Web statistique internationaux 9

2. Comprendre les concepts statistiques 10
Prédiction et recherche 10
Questions à poser aux élèves 10
Types de données 11
Données catégoriques 11
Données numériques 12
Types de diagrammes 13
Diagramme à bandes 13
Diagrammes linéaire (à ligne brisée) 14
Diagrammes circulaires (à secteurs) 15
Diagrammes figuratifs (à pictogramme) 16
Histogrammes 17
Nuages de points 18
Mesures de tendance centrale : moyenne, médiane et mode 19

3. Analyse des données 20
Tâches de gestion des données 20
Utilisation de logiciels 20
Les tableurs 21
Logiciels statistiques dynamiques 21
Logiciel de création de diagrammes dynamiques 22
Logiciels pour calculatrice graphique 22
Lorsqu'on ne dispose d'aucun ordinateur 22
Activités qui nécessitent aucun ordinateur 22

Introduction

Le présent guide vise à soutenir les enseignants de l'élémentaire et du secondaire dans l'enseignement des compétences de base en statistique. Il fournit aux enseignants des instructions précises pour les aider à :

- trouver des ensembles intéressants de données canadiennes qui répondent aux besoins des différents niveaux scolaires;
- choisir les diagrammes convenant à différents types de données;
- calculer des mesures statistiques de base avec l'utilisation ou non de logiciels.

Des compétences essentielles pour les élèves canadiens

Les partenaires de l'éducation au Canada ne mettent plus en doute le fait que nous vivons dans une société de l'information. Autant les enseignants que les parents reconnaissent que les élèves doivent acquérir des compétences dans une foule de domaines techniques pour devenir des citoyens bien instruits et des membres à part entière de la population active. Utiliser l'ordinateur, faire la recherche de données et communiquer en faisant preuve d'un degré élevé de numératie, ce sont effectivement des compétences essentielles aux élèves canadiens.

Les enseignants constatent que plusieurs programmes d'enseignement exigent des élèves des compétences de base en statistique : lecture et création de tableaux, conception de diagrammes et calcul des mesures de tendance centrale (moyenne, médiane et mode). Les élèves mettent ces compétences en pratique non seulement dans les matières qui se fondent traditionnellement sur les chiffres, comme les mathématiques et l'économie, mais également en histoire, en géographie et dans d'autres matières. Les conseils et commissions scolaires partout au Canada favorisent l'acquisition de ces compétences en réservant des périodes de temps importantes aux cours d'informatique au primaire et en offrant de nouveaux cours en gestion de données au secondaire.

De nombreux enseignants s'appuient déjà sur les données de Statistique Canada pour réaliser ces objectifs, plus particulièrement en accédant aux données d'**E-STAT** et du **Recensement à l'école**. Ils constatent que lorsque les élèves utilisent des données canadiennes ou des données au sujet d'eux-mêmes, ils sont plus portés à exercer leur jugement critique au sujet de l'information. Les élèves comparent leurs constatations avec leur propre expérience, remettent en question les données, tirent des conclusions et, en bout de ligne, bâtissent leur propre apprentissage. L'accessibilité grandissante des logiciels de gestion de données permet aux élèves de réaliser plus rapidement l'analyse des données—, l'étape durant laquelle survient la réflexion critique.

La recherche statistique

Les enseignants utilisent les données à différentes fins en classe. Parfois, ils visent principalement à enseigner aux élèves comment travailler avec des données et en tirer des conclusions. À d'autres occasions, les élèves se serviront de données pour analyser certains sujets abordés dans les cours de sciences, de sciences sociales ou d'autres matières. Dans les deux cas, avant d'entreprendre le processus de recherche statistique, il faut deux choses : **une question** nécessitant une réponse et un **ensemble de données.**

Selon le programme étudié, les élèves choisiront d'abord une question et chercheront ensuite des données pour l'appuyer, ou ils feront le contraire.

Nous laisserons à l'enseignant et aux élèves la tâche de définir la question. Cependant, pour leur faciliter la sélection de données pertinentes, la section qui suit présente divers types d'ensembles de données accessibles à partir du site Web de Statistique Canada.

1. Choisir l'ensemble de données

Une recherche statistique requiert un ensemble de données pertinentes. On distingue deux catégories de données : les **données primaires** et les **données secondaires**.

Données primaires

Qu'est-ce qu'on entend par données primaires?

Les données primaires sont recueillies pour répondre à des questions précises. On les obtient directement auprès de sources de première main au moyen d'enquêtes, d'observations ou d'expérimentation.

Les élèves peuvent se servir de **données primaires** provenant d'une enquête prête à utiliser comme le Recensement à l'école (www.recensementecole.ca). Ou bien ils peuvent créer leur propre enquête ou expérience en vue d'obtenir des données sur une question précise. Par exemple, pour répondre à la question « Le temps passé à jouer des jeux d'ordinateur affecte-t-elle les notes scolaires? », les élèves pourraient concevoir une enquête à l'échelle de la classe ou de l'école sur le lien entre les jeux électroniques et les habitudes d'étude.

Ressources pour la collecte de données primaires

Voici les ressources offertes par Statistique Canada pour aider les élèves à recueillir des données primaires :

- **Recensement à l'école**

  www.recensementecole.ca

  Les élèves répondent en classe à une enquête en ligne, en fournissant des renseignements anonymes au sujet de leur vie et de leurs activités. Les questions de l'enquête ont été élaborées par un groupe consultatif interprovincial d'enseignants pour faire en sorte qu'elles correspondent aux programmes d'enseignement et aux intérêts des élèves. Certaines questions, comme les langues parlées à la maison, s'inscrivent dans le programme d'enseignement de la géographie et ressemblent aux questions du recensement national. D'autres questions touchent le programme d'enseignement de la santé, comme l'usage du tabac et la nutrition, ou des sujets sociaux, comme l'intimidation, les modèles de comportement et les oeuvres de charité préférées des élèves.

  Le tableau ci-dessous, qui fait état de certains résultats d'une classe dans le cadre de l'enquête **Recensement à l'école**, constitue un excellent exemple de données primaires.

  Tableau 1
  Exemple de données de classe du Recensement à l'école

Ménage	Langue	Sexe	Âge	Taille	Bras étendus	Yeux	Jeux informatiques	Lecture	Télévision	Pression due au travail scolaire
				centimètres			heures par semaine			
4	le français	masculin	13	172	173	bruns	20	7	9	pression moyenne
3	le français	féminin	14	154	148	bruns	8	0	2	aucune
3	le français	féminin	13	166	161	bleus	0	6	11	pression moyenne
3	le français	masculin	14	166	165	bruns	2	1	3	beaucoup
2	le français	féminin	14	157	150	bruns	4	0	2	beaucoup
6	le français	masculin	14	162	165	autre	15	0	5	pression moyenne
4	le français	féminin	14	170	170	verts	7	4	14	pression moyenne
5	l'anglais	féminin	14	158	157	bleus	7	1	10	pression moyenne

- **Les statistiques : le pouvoir des données!**
  http://www.statcan.ca/francais/edu/power/toc/contents_f.htm
  Ce guide sur les statistiques aidera les élèves à choisir une question particulière, à planifier une enquête et à analyser les données recueillies. Les élèves y trouveront des conseils pratiques sur une foule de sujets, depuis la conception des questionnaires et les méthodes d'échantillonnage jusqu'aux types de diagrammes et les mesures de tendance centrale. Le guide comprend des exercices et des réponses sur de nombreux concepts statistiques, ainsi que des plans de leçons à l'intention des enseignants.

- **Enquête sur les ménages et l'environnement — édition scolaire**
  http://www.statcan.ca/francais/kits/houenv_f.htm
  L'Enquête sur les ménages et l'environnement mesure les pratiques et les comportements environnementaux des Canadiens. La version scolaire permet aux élèves de recueillir leurs propres données au moyen du même questionnaire. Les élèves peuvent ainsi comparer les résultats de leur classe à ceux de l'enquête canadienne officielle.

Données secondaires

Qu'est-ce qu'on entend par données secondaires?

Les élèves peuvent se servir de données recueillies par une autre personne à une autre fin, en autant qu'elles répondent à leurs besoins de recherche. On appelle « **données secondaires** » de tels renseignements obtenus antérieurement.

Les données secondaires peuvent avoir été triées, regroupées en pourcentages, représentées au moyen de diagrammes ou analysées d'une façon quelconque. On peut les obtenir à partir de documents déjà publiés, comme des livres, des revues, des journaux, des publications de recensement gouvernementales ainsi que des rapports ou dossiers d'entreprises.

Sujets pour lesquels les données secondaires s'avèrent utiles

Statistique Canada donne accès en ligne, gratuitement, à toute une gamme de données secondaires canadiennes. Les enseignants trouveront des tableaux de données à l'appui d'une panoplie de sujets de recherche, dont certains pouvant intéresser plus particulièrement les jeunes. Par exemple, les tableaux des indicateurs de la santé (mentionnés ci-dessous dans la section « Sources de données secondaires canadiennes ») fournissent des renseignements sur les questions de santé des jeunes comme :

- les taux de mortalité selon l'âge
- la consommation d'alcool
- la consommation de nourriture
- l'activité physique
- les grossesses chez les adolescentes
- l'usage du tabac

Des élèves qui ont utilisé les ressources de Statistique Canada ont conçu des projets intéressants sur divers sujets, tels que :

- les tendances dans la consommation de l'énergie au Canada
- la grossesse et l'avortement chez les adolescentes
- l'obésité et le diabète : une épidémie grandissante
- les infractions routières des adolescents
- l'éducation, le salaire et les cheminements de carrière

Sources de données secondaires canadiennes

Les ressources suivantes serviront à trouver des sujets qui correspondent aux programmes d'enseignement et aux intérêts des élèves.

- **E-STAT**

  http://estat.statcan.ca

  En mathématiques et en sciences sociales, vous pouvez avoir recours à E-STAT pour produire des diagrammes, créer des tableaux, et télécharger, analyser et cartographier une multitude de données. Il s'agit de la source de données la plus détaillée de Statistique Canada; elle couvre l'écart d'années le plus vaste et les niveaux géographiques les plus détaillés. E-STAT contient des données de séries chronologiques qui relèvent de centaines d'enquêtes socio-économiques, ainsi que plusieurs centaines de variables du recensement pour chaque province, comté et municipalité au Canada, allant jusqu'aux quartiers (qu'on nomme secteurs de recensement). Notons l'intérêt particulier des données sur les comportements de santé des enfants d'âge scolaire. Pour accéder à E-STAT, l'inscription sans frais de l'école est exigée.

- **Profils des communautés**

  http://www12.statcan.ca/english/census06/data/profiles/community/Index.cfm?Lang=F

  Visitez votre localité! Au fur et à mesure que les données du Recensement de 2006 deviennent disponibles, on les ajoute aux *Profils des Communautés de 2006* pour toutes les communautés au Canada. Ces données portent sur cinq sujets principaux : la population; les gains et le revenu; la scolarité; le travail; les familles et les logements. Vous pouvez obtenir des données de 2001 sur tous ces sujets en consultant les *Profils des communautés de 2001* qui contiennent plus de 150 variables provenant du Recensement de la population, du Recensement de l'agriculture et de l'Enquête auprès des peuples autochtones.

- **Tableaux sommaires**

  http://www40.statcan.ca/index_f.htm

  Les Tableaux sommaires, auparavant appelés *Le Canada en statistiques*, fournissent des données nationales et provinciales constamment mises à jour qui relèvent de centaines d'enquêtes. Ces 500 tableaux sélectionnés sont répartis en 32 grandes catégories de sujets. On peut les consulter selon le sujet, la province, le territoire ou la région métropolitaine. La section « Aperçu » permet la consultation d'articles courts et utiles sur chaque sujet.

- **Indicateurs de la santé**

  http://www.statcan.ca/bsolc/francais/bsolc?catno=82-221-X

  Plus de 80 indicateurs permettent de mesurer la santé de la population canadienne et l'efficacité du système de soins de santé. Conçues pour établir des comparaisons au niveau des régions sociosanitaires et des provinces et territoires, ces données sont produites à partir d'un large éventail de sources des plus récentes.

- **Données du recensement**

  http://www12.statcan.ca/francais/census/index.cfm

  E-STAT et les Profils des communautés sont les principales sources pour les élèves qui cherchent des données du recensement, mais la section « Recensement » du site Web de Statistique Canada fournit des ressources supplémentaires. On y trouve notamment des renseignements au sujet du recensement, des articles de la série « Analyses », des tableaux de données organisés par sujet et par niveau géographique, des cartes, des exposés animés et des documents de référence.

- ***Le Quotidien***

  http://www.statcan.ca/francais/dai-quo/

  Les communiqués du bulletin *Le Quotidien* portent sur une multitude de sujets et comportent généralement des tableaux ou des diagrammes. Utilisez le moteur de recherche « Recherche dans Le Quotidien » pour trouver des articles d'intérêt, puis cliquez sur « classement par date » pour trouver les articles les plus récents. Les articles font souvent référence aux numéros de tableaux CANSIM, lesquels vous permettent d'examiner les données de façon plus approfondie. On peut accéder gratuitement à ces tableaux dans E-STAT. Les articles contiennent aussi un lien avec la section « Définitions, sources de données et méthodes : numéro d'enquête », où il est possible de se renseigner au sujet des techniques d'échantillonnage et de lire le questionnaire d'enquête.

- **Publications**

  http://www.statcan.ca/francais/pub/index_f.htm

  Cette section fournit des listes de publications : les publications en format électronique sont offertes gratuitement, alors que celles sous forme d'imprimées sont à vendre. On peut trouver des tableaux de données propres à des sujets particuliers dans les publications qui traitent de l'environnement, de la santé, de l'éducation, de la justice et d'autres sujets.

- **Fichiers de microdonnées**

  http://www.statcan.ca/francais/kits/courses/math2_f.htm

  Les fichiers de microdonnées constituent une excellente source d'ensembles de données volumineux se prêtant à la manipulation au moyen de logiciels statistiques dynamiques (p.ex. Fathom). Ils fournissent un échantillon représentatif de réponses individuelles anonymes qui proviennent d'enquêtes. Pour des exemples, voir les microdonnées relatives aux enquêtes suivantes :
  - le Recensement de 2001
  - l'Enquête longitudinale nationale sur les enfants et les jeunes

Sources de données secondaires internationales

- **Données internationales du Recensement à l'école**

  http://www.censusatschool.org/international/index.asp?country=&lang=fr

  En plus de fournir des données primaires au sujet de votre propre classe, le **Recensement à l'école** permet aux élèves d'accéder aux données secondaires recueillies dans un certain nombre de pays, notamment le Royaume-Uni, le Canada, la Nouvelle-Zélande, l'Australie et l'Afrique du Sud. Ces données sont utiles pour répondre aux questions comparant des situations à une échelle internationale.

- **Données internationales de E-STAT**

  http://estat.statcan.ca/content/francais/hbsc/hbsc_start_f.shtml

  La plupart des données de E-STAT sont canadiennes mais on y trouve aussi certains renseignements provenant de d'autres pays. Les élèves s'intéressent particulièrement aux données de différents pays sur les comportements de santé des jeunes d'âge scolaire.

- **Sites Web statistiques internationaux**

  http://www.statcan.ca/francais/reference/othsit_f.htm

  Les élèves peuvent cliquer sur ces liens pour accéder aux bureaux de statistique nationaux et aux autres organisations de statistique internationales et pour examiner les tendances existantes dans des dizaines de pays aux fins de la réalisation de projets sur des questions mondiales.

2. Comprendre les concepts statistiques

Prédiction et recherche

Tel que mentionné dans l'introduction, les recherches statistiques débutent parfois par une simple question d'un élève, laquelle amène la classe à recueillir les données primaires qui pourront y répondre. D'autres fois, on dispose d'un ensemble de données secondaires qui peut mener les élèves à se poser des questions intéressantes.

Dans un cas comme dans l'autre, les élèves comprennent mieux les données s'ils sont amenés à réfléchir au sujet des réponses possibles avant, pendant et après la collecte des données. Les élèves doivent apprendre à **prédire** la réponse et à **faire une recherche** pour la vérifier.

Une fois la question posée, les élèves doivent commencer à réfléchir à la façon dont ils peuvent trouver la réponse. On peut commencer en demandant aux élèves de **prédire** ou de **supposer** ce que la réponse sera selon eux, puis à déterminer de quelle façon ces hypothèses mènent à une **recherche** sur le problème.

Considérons la question de recherche suivante : Le temps consacré à jouer à des jeux d'ordinateur affecte-t-il les notes scolaires?

Les élèves vont tendre à se regrouper d'abord en un camp du « non » et en un camp du « oui ». En leur demandant d'expliquer leurs réponses, vous les encouragerez à réfléchir aux conditions qu'ils rattacheraient à leur réponse.

Ils peuvent par exemple avancer que la réponse dépend de l'âge ou du sexe des élèves. Ou encore, ils diront que le temps consacré aux leçons et aux devoirs a lui aussi un effet sur les notes, tout comme celui consacré aux jeux d'ordinateur. Cette idée peut les amener à inclure dans leur enquête des questions sur les leçons et les devoirs en plus de celles sur les jeux d'ordinateur. Ainsi, en réfléchissant à leurs prédictions initiales, ils en déduiront les étapes de la recherche.

De plus, lorsque vous demandez aux élèves de prévoir la réponse avant de faire leur recherche, vous les aidez à s'apercevoir des idées erronées et à les corriger au moment de leur collecte de données.

Questions à poser aux élèves

- Que pensez-vous sera la réponse à votre question?
- Prévoyez-vous des exceptions à votre prévision?
- Quelles données recueillerez-vous?
- De quelle façon recueillerez-vous les données?
- Qui interrogerez-vous?
- Quel système de mesure utiliserez-vous?
- De quelle façon consignerez-vous l'information?

Types de données

Les questions particulières produisent des types particuliers de données, lesquelles à leur tour se prêtent à un type particulier de diagramme.

On distingue deux principaux types de données : les données **catégoriques** et les données **numériques**.

Données catégoriques

La question « Quelle couleur sont vos cheveux? » produit des données catégoriques qui se classent parmi les catégories « bruns », « blonds », « noirs », « roux » ou « d'une autre couleur ». Les données catégoriques peuvent être subdivisées en deux sous-types : les données nominales ou ordinales.

Le tableau ci-dessous présente chaque sous-type et les types de diagramme qui y sont associés.

Tableau 2
Données catégoriques

Types de données	Sous-types	Exemples de la base de données Recensement à l'école	Diagrammes pertinents
Catégoriques : Les données se classent dans diverses catégories de réponses à une question.	**Données nominales :** Elles sont désignées au moyen de noms particuliers ou de catégories particulières. Ces données ne peuvent être organisées dans aucun autre ordre « naturel ».	**Sexe :** masculin ou féminin **Matières préférées :** mathématiques, histoire, éducation physique, musique, etc. **Couleur des yeux :** bruns, bleus, verts ou autre couleur **Animaux de compagnie :** chats, chiens, oiseaux, poissons, etc.	Diagramme à bandes, diagramme circulaire, pictogramme.
	Données ordinales : Elles sont désignées au moyen de catégories pouvant être classées dans un ordre particulier ou d'une certaine « façon naturelle ».	**Pression due au travail scolaire :** aucune, très peu, pression moyenne, beaucoup	Diagramme à bandes, diagramme circulaire pictogramme

Données numériques

La question « Combien de personnes habitent dans votre ménage? » produit des données numériques. Celles-ci peuvent être subdivisées en deux sous-types : les données **discrètes** ou **continues**.

Le tableau ci-dessous présente chaque sous-type et les types de diagramme qui y sont associés.

Tableau 3
Données numériques

Types de données	Sous-types	Exemples de la base de données Recensement à l'école	Diagrammes pertinents
Données numériques : Elles sont représentées par des nombres réels. On les appelle également données **quantitatives**.	**Données discrètes :** Elles comprennent un nombre déterminé de réponses différentes. Par exemple, le nombre de personnes faisant partie d'un ménage représente une donnée discrète parce qu'on peut seulement répondre au moyen d'un nombre entier. On ne peut inclure comme réponses possibles toutes les décimales ou les fractions qui se situent entre ces nombres entiers. Par exemple, un ménage ne peut pas compter 2,5 ou 3,75 personnes.	**Âge en années :** 7, 8, 9, 10, 11, etc.	Diagramme à bandes, diagramme linéaire, diagramme circulaire, histogramme.
Nota – Parfois les nombres correspondent à des échelles de réponses (p. ex. : 0 = pas du tout, 1 = très peu, 2 = quelques-uns, etc.). On considère alors ces réponses comme des données catégoriques ordinales plutôt que comme des données numériques, même si elles sont représentées par un nombre.	**Données continues :** Elles peuvent comprendre un nombre infini de réponses différentes. Les réponses possibles sont infinies puisqu'elles peuvent inclure des valeurs décimales. Par exemple, la taille d'un élève peut être de 1,57923 mètres.	**Nombre de personnes faisant partie d'un ménage :** 1, 2, 3, 4, 5, etc. **Nombre de jours durant lesquels on s'est adonné à une activité physique intense la semaine dernière :** 0, 1, 2, 3, 4, 5, 6, 7, etc. **Taille, dimension des bras étendus, circonférence du poignet :** Il est impossible d'énumérer toutes les possibilités. **Nota –** Dans l'enquête Recensement à l'école, les élèves doivent arrondir leurs réponses au centimètre ou au millimètre près, de sorte que leurs réponses constituent en fait des données discrètes.	Diagramme linéaire, histogramme.

Nota : Pour faciliter le traitement des **données continues**, on les regroupe souvent en « intervalles de classes ». Le regroupement des données fait partie du processus d'organisation des données afin de rendre l'information utile. Par exemple, au lieu de représenter toutes les tailles d'élèves mesurées dans une classe, il est plus efficace de représenter des catégories regroupées telles que : 120 à 129 cm, 130 à 139 cm, 140 à 149 cm, etc.

Les données discrètes peuvent être regroupées ou non. Le regroupement des données facilite leur traitement, mais lorsqu'il s'agit d'un nombre restreint de réponses, on peut aussi bien les laisser non regroupées.

Types de diagrammes[1]

Diagrammes à bandes

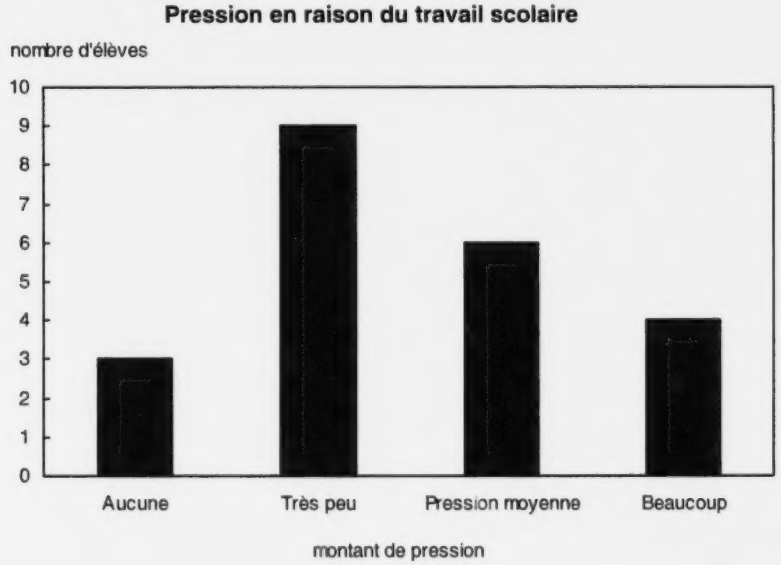
Les diagrammes à bandes présentent des données **catégoriques** ou **numériques.** Les données numériques peuvent être **non regroupées** (si elles comprennent un nombre restreint de réponses) ou être **regroupées** en intervalles de classes.

Les diagrammes à bandes sont constitués d'un axe comportant des bandes horizontales ou verticales munies d'étiquettes. Les diagrammes à bandes verticales sont également appelés diagrammes à colonnes. Les bandes illustrent les fréquences des différentes réponses. Les nombres figurant sur l'axe des x d'un diagramme à bandes horizontales ou sur l'axe des y d'un diagramme à bandes verticales sont appelés l'**échelle**.

Lorsqu'on crée des diagrammes à bandes, on représente chaque catégorie ou valeur au moyen d'une bande verticale ou horizontale. La hauteur ou la longueur de la bande représente le nombre d'unités ou d'observations dans cette catégorie (c.-à-d. leur fréquence).

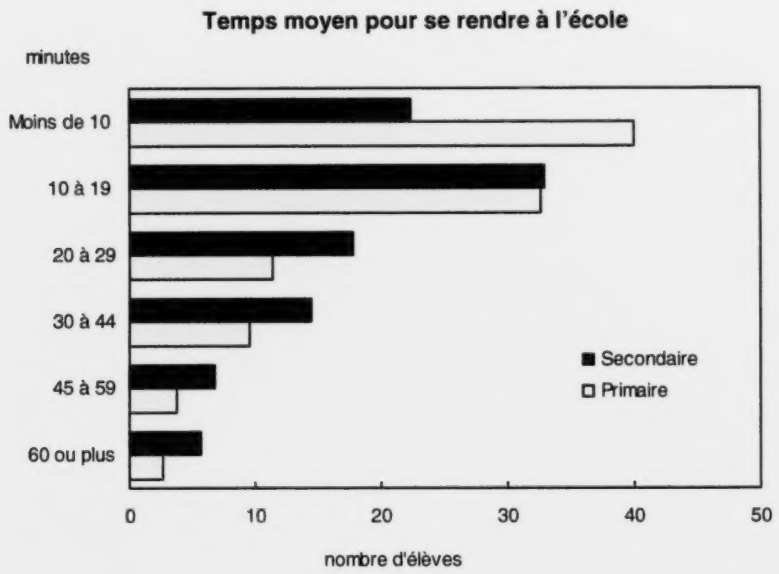
Il est conseillé de ne pas créer des diagrammes à bandes en trois dimensions, car l'ajout de la profondeur peut rendre plus difficile la lecture des données.

Figure 1 Exemple de diagramme à bandes verticales :



1. Pour plus de renseignements, voir *Les statistiques : le pouvoir des données*, Types de graphiques http://www.statcan.ca/francais/edu/power/ch9/first9_f.htm

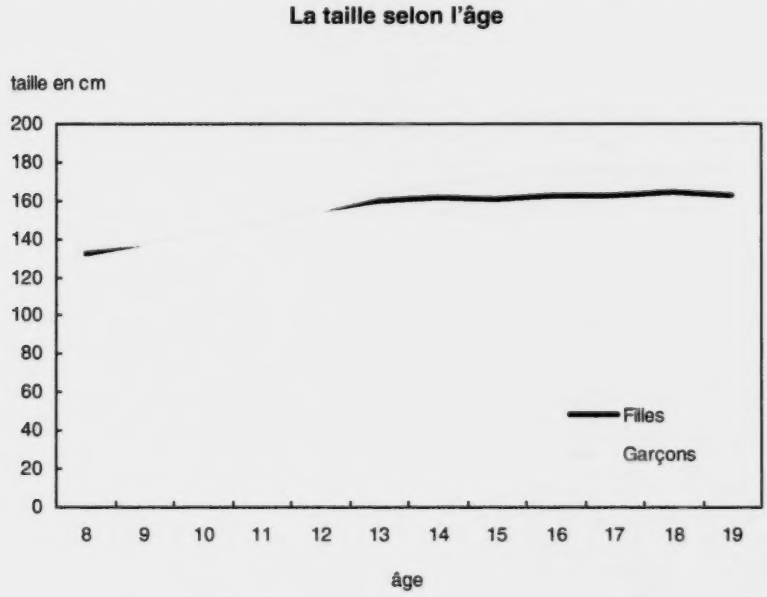
Figure 2
Exemple de diagramme à bandes horizontales :



Diagrammes linéaire (à ligne brisée)

Les diagrammes linéaires ou à ligne brisée comparent deux variables : l'une est placée sur l'axe des x (horizontal) et l'autre, sur l'axe des y (vertical). Le diagramme montre le rapport entre les variables ou les variations entre elles au moyen d'un trait continu entre tous les points.

Figure 3 Exemple de diagramme linéaire illustrant le rapport entre deux variables :

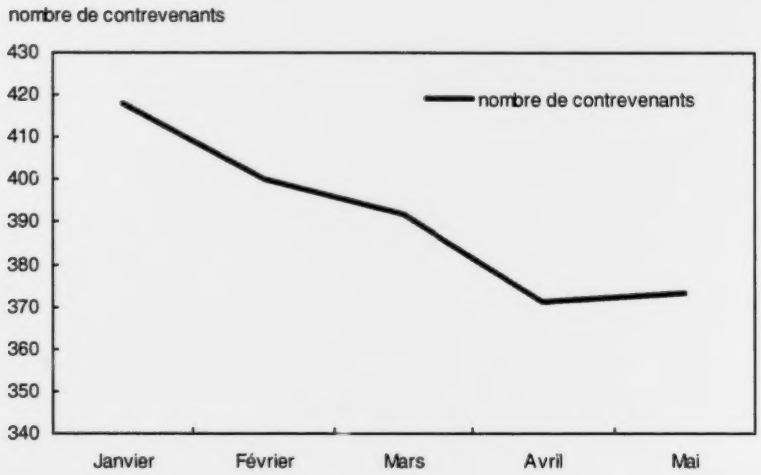


Les diagrammes linéaires servent également à révéler les tendances dans le temps. Alors que les diagrammes à bandes révèlent une variation de l'ampleur, les diagrammes linéaires indiquent un changement d'orientation. Ces dernières sont populaires en vue de montrer la variation des données dans le temps parce qu'elles révèlent clairement les tendances tout en étant faciles à créer.

Lorsqu'un diagramme linéaire illustre une tendance dans le temps, l'axe des y indique habituellement la quantité (p. ex. dollars, litres) ou le pourcentage, tandis que l'axe des x mesure les unités de temps.

Figure 4 Exemple de diagramme linéaire illustrant une tendance dans le temps :

Nombre de contrevenants reconnus coupables, Grishamville



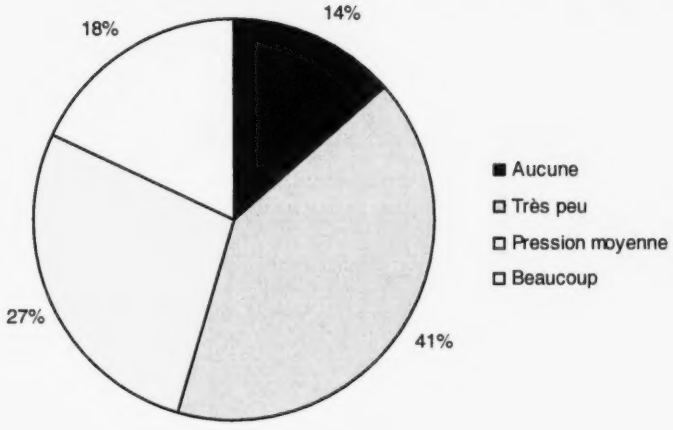
Diagrammes circulaires (à secteurs)

Un diagramme circulaire ou à secteurs constitue une façon de résumer un ensemble de **données catégoriques** ou de présenter les différentes valeurs d'une variable donnée (p. ex. la répartition en pourcentage). Ce type de diagramme est formé d'un cercle subdivisé en segments, chaque segment représentant une catégorie particulière et son pourcentage de l'ensemble. La surface de chaque segment représente une proportion du cercle qui est identique à la proportion de la catégorie par rapport à l'ensemble des données.

Il est conseillé d'utiliser les diagrammes circulaires lorsque le nombre de catégories est restreint – idéalement pas plus de six – sans quoi le tableau obtenu sera trop difficile à comprendre. N'utilisez jamais un diagramme circulaire tridimensionnel, même lorsqu'un logiciel de tableur l'offre comme option de diagramme. L'image tridimensionnelle est trompeuse parce que la surface de certains segments peut sembler plus importante que les proportions réelles qu'ils représentent.

Figure 5 Exemple de diagramme circulaire :

Pourcentage d'élèves



pourcentage d'élèves

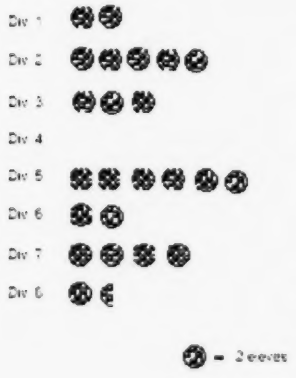
Diagrammes figuratifs (à pictogramme)

Un diagramme figuratif ou à pictogramme emploie des symboles pour faire ressortir la signification des **données catégoriques.** Il ressemble à un diagramme à bandes du fait que chaque rangée horizontale ou verticale représente la fréquence ou le nombre de réponses dans chaque catégorie. Il est conseillé d'utiliser les diagrammes figuratifs avec prudence parce que les figures peuvent, accidentellement ou volontairement, donner une fausse représentation des données.

Par exemple, le symbole du biscuit dans le diagramme figuratif ci-dessous représente deux élèves et le demi-biscuit en représente un. D'autres types de diagrammes figuratifs peuvent utiliser un symbole dont les dimensions augmentent ou diminuent pour représenter les variations des données. En pareil cas, il faut veiller à ce que les dimensions ou l'étendue (surface totale) du symbole soient proportionnelles aux variations qu'il représente.

Figure 6 Exemple de diagramme figuratif :

Nombre d'élèves qui préfèrent les biscuits aux brisures de chocolat



Histogrammes

On utilise un histogramme pour résumer des données numériques **continues** ou **discrètes** mesurées dans une échelle d'intervalle. On y a souvent recours pour illustrer les caractéristiques principales de la répartition des données. L'histogramme subdivise la fourchette des valeurs possibles en catégories ou en groupes. On construit, pour chaque groupe, un rectangle dont la longueur de la base correspond à la fourchette des valeurs du groupe en question et dont la hauteur est proportionnelle au nombre d'observations ou à la fréquence du groupe. En d'autres termes, les rectangles auront des hauteurs différentes. L'histogramme a un aspect semblable à un diagramme à bandes verticales, mais lorsque les variables sont continues, il n'existe aucun écart entre les bandes. Toutefois, lorsque les variables sont discrètes, on devrait laisser des écarts entre les bandes.

Figure 7
Exemple d'histogramme illustrant des données continues :

Distribution du nombre d'élèves d'une classe de 6e année selon la taille

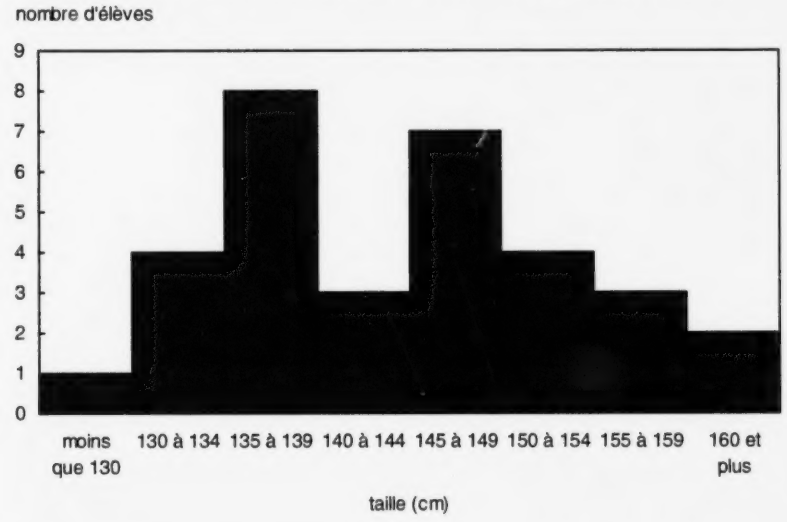
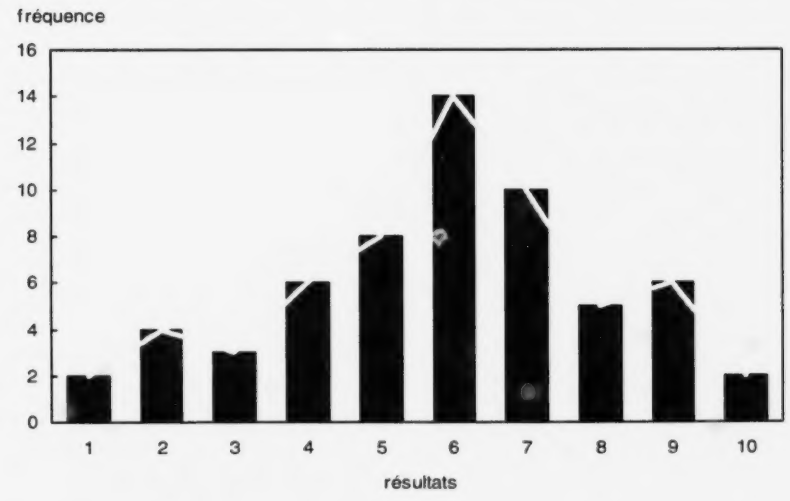


Figure 8 Exemple d'histogramme illustrant des données discrètes :

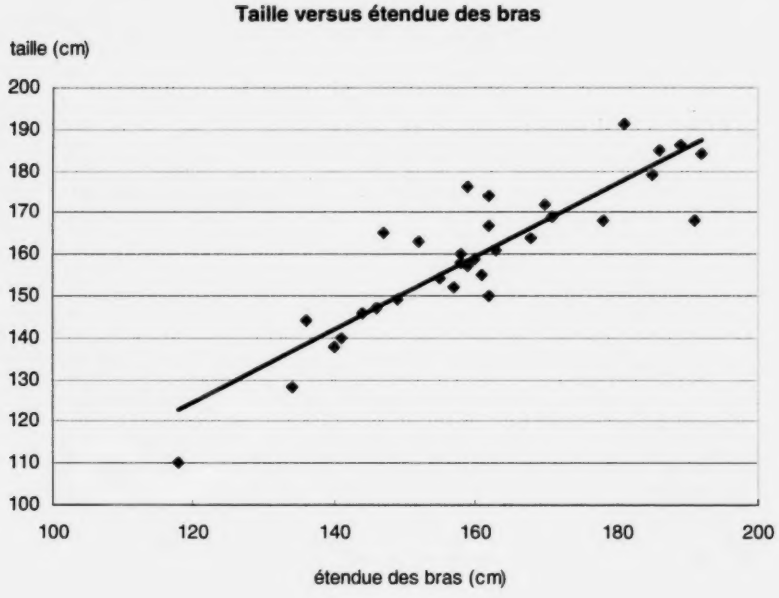
Fréquence des résultats d'un test de mathématique de 10 questions



Nuages de points

On utilise les nuages de points pour représenter la relation entre deux variables au moyen de paires ordonnées tracées sur un plan cartésien. Les points ne sont pas reliés; l'image qui en résulte indique le type et l'intensité de la relation entre les variables. On peut tracer une droite de meilleur ajustement entre les points lorsqu'il existe une relation. Les nuages de points peuvent illustrer les notions suivantes : la corrélation statistique, la relation positive ou négative entre les variables, les tendances non linéaires, la répartition des données et les valeurs aberrantes.

Figure 9 Exemple de nuage de points illustrant une corrélation positive :



Mesures de tendance centrale : moyenne, médiane et mode[2]

La moyenne, la médiane et le mode servent à dégager les caractéristiques d'un ensemble de données au moyen d'un seul nombre. Par exemple, une classe typique de Secondaire I (7e année) pourrait être composée d'élèves de 12 à 15 ans. Cependant, comme la classe compte plus d'élèves de 14 ans que de tout autre groupe d'âge, nous utilisons l'âge modal de 14 ans pour représenter l'âge des élèves de Secondaire I. Selon la situation, la meilleure description d'un ensemble de données particulier pourrait être soit la moyenne, la médiane ou le mode.

On calcule la **moyenne** en additionnant toutes les données et en divisant la somme par le nombre total de données.

La **médiane** est le nombre **du milieu** lorsque les données de l'ensemble ont été classées en ordre numérique : exactement la moitié des données se situent alors au-dessus de la médiane et la moitié se situent au-dessous. Il faut donc trier les données dans un ordre ascendant ou descendant pour déterminer le nombre du milieu. Si l'ensemble compte un nombre pair de données, il faut calculer la moyenne des deux nombres du milieu pour trouver la médiane.

Le **mode** est le nombre observé **le plus fréquemment** dans l'ensemble. Lorsque deux nombres sont observés aussi souvent l'un et l'autre, les données sont bimodales.

Dans une **distribution normale**, la moyenne, la médiane et le mode ont des valeurs identiques. Par exemple, l'ensemble de données qui suit montre une distribution normale :

Ensemble de données : 14, 14, 13, 15, 15, 14, 13, 14, 13, 15

Moyenne : (14 + 14 + 13 + 15 + 15 + 14 + 13 + 14 + 13 + 15) / 9 = 14

Médiane : la médiane est 14 (13, 13, 13, 14, **14**, **14**, 14, 15, 15, 15)

Mode : le nombre observé le plus fréquemment est 14

2. Pour plus de renseignements, voir *Les statistiques : le pouvoir des données*, Mesures de tendance centrale, http://www.statcan.ca/francais/edu/power/ch11/first11_f.htm

3. Analyse des données

Tâches de gestion des données

Une fois qu'on a obtenu un ensemble de données, le moment est venu d'en commencer l'analyse. Il s'agit donc de réaliser une partie ou la totalité des activités suivantes :

- l'organisation des données en tableaux
- le calcul des pourcentages
- le calcul des mesures de tendance centrale
- la construction de diagrammes

En plus d'exécuter ces tâches mécaniques, les élèves doivent réfléchir à la signification des données.

Ils peuvent par exemple étudier le rapport entre deux variables ou observer l'effet de l'élimination des valeurs extrêmes sur la moyenne ou la médiane. La plupart des programmes d'enseignement provinciaux de la 7e et 8e année (Secondaire I et II) prévoient que les élèves développent leurs compétences dans le domaine de la gestion des données en apprenant notamment à

- cueillir et organiser des données catégoriques, discrètes et continues
- présenter des données au moyen de tableaux de fréquences relatives, de diagrammes à bandes, de diagrammes circulaires, d'histogrammes, etc.
- identifier le biais dans les données
- établir un lien entre les variations des données et les variations de la tendance centrale
- formuler des conclusions à partir de données

Utilisation de logiciels

Grâce à de nombreux logiciels, on peut effectuer la plus grande partie du travail mécanique de création des tableaux, d'exécution des calculs et de construction des diagrammes au moyen de quelques clics de la souris. En éliminant le travail lent et fastidieux qui peut constituer un obstacle pour les élèves, ces logiciels leur permettent de progresser rapidement à l'étape où ils dégageront un sens des données.

Les élèves peuvent utiliser un logiciel pour interroger les données et expérimenter différents types de diagrammes ou pour examiner les rapports entre des variables. Ainsi, dès la quatrième année, les élèves peuvent dégager un sens des données au cours d'une seule période de classe sans avoir à passer toute la période à dessiner un seul diagramme. De plus, comme la majorité des élèves aiment utiliser les outils électroniques, ils se divertissent tout en travaillant.

Si vous disposez d'un accès limité aux ordinateurs, rappelez-vous qu'une des meilleures méthodes d'explorer les données est le travail en petits groupes, ce qui permet aux élèves de se poser des questions du genre « Je me demande si... », puis de les vérifier ensemble. Il n'est pas nécessaire de disposer d'un ordinateur par élève puisque de nombreuses activités de gestion de données se prêtent à l'apprentissage en groupe de deux ou quatre. Quatre à six ordinateurs peuvent ainsi suffire pour donner un cours.

Les divers conseils et commissions scolaires au Canada utilisent de nombreux logiciels. La liste ci-dessous offre un aperçu des programmes les plus couramment utilisés, mais elle ne se veut pas exhaustive. Veuillez noter que **Statistique Canada** n'appuie aucunement l'utilisation d'un programme particulier.

On peut classer les logiciels de gestion de données en sous-groupes :

- les tableurs
- les logiciels statistiques dynamiques
- les logiciels de création de diagrammes dynamiques
- les logiciels pour calculatrice graphique

Les tableurs

Exemples :

- Excel (4e à la 12e année/cégep)
  http://office.microsoft.com/fr-ca

- QuattroPro (4e à la 12e année/cégep)
  http://www.functionx.com/quattropro/lesson01.htm
  Nota : site en anglais seulement

- AppleWorks (4e à la 12e année/cégep)

http://download.info.apple.com/Apple_Support_Area/Manuals/software/FUManueldelutilisateurAW.PDF

- Calc dans Open Office (4e à la 12e année/cégep)
  http://www.framasoft.net/article3640.html

Les données importées dans un tableur sont présentées sous forme d'un tableau constitué de rangées et de colonnes. On peut calculer rapidement les fourchettes de données, les totaux de colonnes, les moyennes et les pourcentages en inscrivant des formules dans les cellules pertinentes. Certains programmes offrent des options pour certains calculs à partir d'un seul clic.

Pour représenter les données sous forme de diagramme, utilisez l'outil de création de diagrammes de votre tableur. Lorsqu'il s'agit de données catégoriques, on doit d'abord produire un tableau de fréquence avant de créer le diagramme. De plus, Excel offre l'option **rapport de tableaux dynamiques**.

L'outil de création de diagrammes permet aux élèves de sélectionner un type de diagramme et de le personnaliser au moyen de couleurs, de polices de caractères, etc. Voir **Types de données** à la page 7 du présent document pour identifier les types de diagrammes qui conviennent à différents types de données.

Logiciels statistiques dynamiques

Exemples :

- TinkerPlots (4e à la 8e année/Secondaire II)
  http://www.keypress.com/x5715.xml
  Nota : site en anglais seulement

- Fathom (8e à la 12e année/Secondaire II à cégep)
  http://www.keypress.com/x5656.xml
  Nota : site en anglais seulement

On importe les données dans ces programmes en coupant-collant une adresse URL ou en important un fichier de tableur dans le format « champs séparées par des virgules » (.csv) ou « texte simple » (.txt). Les données apparaissent sous forme d'une liste d'attributs dans une carte de données. On peut transférer la liste d'attributs dans un diagramme par glisser-déplacer. Chaque point figurant dans le diagramme représente la réponse d'une personne de l'échantillon. Il est possible d'afficher l'ensemble complet d'attributs de la personne en question en cliquant sur le point.

Les élèves peuvent poser une question, puis créer un ou deux diagrammes différents simplement en faisant glisser des colonnes de données sur les axes. Ils peuvent poser une question différente, puis faire glisser les données d'une autre colonne et observer immédiatement les changements.

Voici quelques-uns des types de diagrammes offerts dans ces logiciels :

- le diagramme de points (à échelle superposée ou non superposée)
- le diagramme linéaire
- l'histogramme
- le diagramme à cases
- le diagramme à bandes

- le nuage de points

Logiciel de création de diagrammes dynamiques

Exemple :

- Graphe Easy (8e à la 12e année/Secondaire II à cégep)
  http://www.grapheeasy.com/

Les logiciels de création de diagrammes sont utiles pour étudier les rapports entre deux variables. On introduit les données point par point dans une grille cartésienne (axes des x et y) pour créer un nuage de points. Les fonctionnalités du logiciel offrent la possibilité de déterminer la droite de meilleur ajustement parmi les points et d'obtenir un sommaire instantané des mesures statistiques (moyenne, variance et écart type). Les élèves peuvent examiner l'effet que l'élimination des valeurs aberrantes peut avoir sur la forme du diagramme ou sur la moyenne.

Logiciels pour calculatrice graphique

Exemple :

- TI-Connect destiné à l'utilisation avec la calculatrice graphique TI
  (7e à la 12e année/Secondaire I à V et cégep)
  http://education.ti.com/educationportal/sites/FRANCE/homepage/index.html

Les logiciels pour calculatrice graphique permettent aux élèves d'importer des données dans leur calculatrice afin d'étudier les rapports entre deux variables. Les élèves peuvent aussi introduire manuellement les données de chaque variable dans un tableau de valeurs de la calculatrice.

La calculatrice graphique peut représenter les données sous forme d'un nuage de points; elle fournira les droites de meilleur ajustement et les mesures statistiques pertinentes (moyenne, variance et écart type). La plupart des programmes d'enseignement provinciaux exigent l'utilisation de calculatrices graphiques dans le cadre de certains cours de mathématiques de la 8e à la 11e année (Secondaire II à V).

Lorsqu'on ne dispose d'aucun ordinateur

Comme plusieurs écoles ne disposent pas toujours d'ordinateurs et de logiciels, une grande partie du travail doit se faire manuellement. Même lorsqu'on a facilement accès à des ordinateurs, il peut être souhaitable pour les élèves d'effectuer manuellement le travail de certaines parties du programme d'enseignement. En réalisant eux-mêmes les étapes manuelles, les élèves peuvent mieux comprendre les concepts inhérents à la création de tableaux et de diagrammes et au calcul de mesures statistiques.

Il existe de nombreuses activités qui permettent de travailler avec les données « manuellement ». La liste qui suit présente certaines de ces activités, qui peuvent être imprimées ou téléchargées à partir d'Internet. Il est à noter que dans certains cas, avant de commencer l'activité, il faut imprimer des données choisies d'un tableur afin d'en fournir une copie papier aux élèves.

Activités qui nécessitent aucun ordinateur

- **Le recherchiste, c'est vous!** (4e à la 8e année/Secondaire II)
  http://www19.statcan.ca/02/02_037_f.htm
  Ce mini-projet, réalisé au moyen de données du Recensement à l'école, demande aux élèves de choisir une question intéressante à étudier à l'aide des résultats de leur classe. Ils présenteront ensuite les résultats au moyen de différents types de diagrammes.

- **Feuilles d'activités : analyse des données de la classe**
  (9e à la 12e année/Secondaire III à cégep)
  http://www19.statcan.ca/02/02_046_f.htm
  Ces activités utilisent les données de classe du Recensement à l'école pour couvrir les mesures de tendance centrale, les tableaux de fréquences, les diagrammes circulaires et les histogrammes, les droites de meilleur ajustement, etc. Une rubrique d'évaluation est comprise.
  Nota : Dans la dernière activité, on demande aux élèves de comparer leur droite de meilleur ajustement (faite à la main) avec une droite créée au moyen d'une calculatrice graphique. On peut omettre cette partie si on ne dispose pas de calculatrice graphique.

- **Autres activités pédagogiques du Recensement à l'école**
  (4e à la 12e année/cégep)
  http://www19.statcan.ca/02/02_004_f.htm
  Ces activités utilisent les résultats de classe du Recensement à l'école dans les programmes d'enseignement des mathématiques, des sciences sociales et de la santé. Même si certaines leçons demandent aux élèves d'extraire des données canadiennes ou internationales du site Web au moyen d'ordinateurs, la plupart d'entre elles exige seulement que l'enseignant imprime des colonnes sélectionnées de données de la classe avant l'activité.

  Exemples de titres de leçons :

  En route pour l'école : (4e à la 8e année/Secondaire II) Analyse des données au moyen de tableaux « tiges et feuilles » et de diagrammes circulaires.

  Quel zoo! : (4e à la 8e année/Secondaire II) Examen de diagrammes à bandes représentant les mêmes données mais utilisant des échelles différentes.

  La théorie de Vitruve s'applique-t-elle à vous? : (4e à la 8e année/Secondaire II) et (9e à la 12e année/cégep) Vérification de la théorie illustrée par Léonard de Vinci selon laquelle la dimension des bras étendus est égale à la taille.

- **Plans de leçons E-STAT** (7e à la 12e année/Secondaire I au cégep)
  http://www.statcan.ca/francais/Estat/lesson_f.htm
  Ces activités utilisent des données d'E-STAT dans les programmes d'enseignement des mathématiques, des sciences sociales et de la santé. Des instructions et des feuilles de travail complètes destinées aux élèves sont incluses. Même si la plupart des leçons de cette section demandent aux élèves d'utiliser des ordinateurs pour extraire les tableaux de données d'E-STAT, l'enseignant peut facilement imprimer les tableaux et les distribuer aux élèves avant l'activité.

  Exemples de titres de leçons :

  Comparaison des choix alimentaires et de l'image de soi des jeunes de 15 ans partout dans le monde (7e à la 12e année/Secondaire I au cégep) :
  Examen et comparaison des données sur les jeunes de différents pays.

  Jeu de rôle : interprétation de Jean Talon (7e à la 9e année/Secondaire I à III) :
  En jouant le rôle de Jean Talon, utiliser les données du Recensement de 1665 pour convaincre le roi de France d'investir davantage dans la Nouvelle-France.

  Fumer aujourd'hui = fumer demain (7e à la 9e année/Secondaire I à III) :
  Recherche sur les tendances en examinant les données sur le tabagisme selon l'âge et le sexe.